

Collective Animal Behavior from Bayesian Estimation and Probability Matching

Alfonso Pérez-Escudero*, Gonzalo G. de Polavieja*

Abstract

Animals living in groups make movement decisions that depend, among other factors, on social interactions with other group members. Our present understanding of social rules in animal collectives is mainly based on empirical fits to observations, with less emphasis in obtaining first-principles approaches that allow their derivation. Here we show that patterns of collective decisions can be derived from the basic ability of animals to make probabilistic estimations in the presence of uncertainty. We build a decision-making model with two stages: Bayesian estimation and probabilistic matching. In the first stage, each animal makes a Bayesian estimation of which behavior is best to perform taking into account personal information about the environment and social information collected by observing the behaviors of other animals. In the probability matching stage, each animal chooses a behavior with a probability equal to the Bayesian-estimated probability that this behavior is the most appropriate one. This model derives very simple rules of interaction in animal collectives that depend only on two types of reliability parameters, one that each animal assigns to the other animals and another given by the quality of the non-social information. We test our model by obtaining theoretically a rich set of observed collective patterns of decisions in three-spined sticklebacks, *Gasterosteus*

aculeatus, a shoaling fish species. The quantitative link shown between probabilistic estimation and collective rules of behavior allows a better contact with other fields such as foraging, mate selection, neurobiology and psychology, and gives predictions for experiments directly testing the relationship between estimation and collective behavior.

Author Summary

Animals need to act on uncertain data and with limited cognitive abilities to survive. It is well known that our sensory and sensorimotor processing uses probabilistic estimation as a means to counteract these limitations. Indeed, the way animals learn, forage or select mates is well explained by probabilistic estimation. Social animals have an interesting new opportunity since the behavior of other members of the group provides a continuous flow of indirect information about the environment. This information can be used to improve their estimations of environmental factors. Here we show that this simple idea can derive basic interaction rules that animals use for decisions in social contexts. In particular, we show that the patterns of choice of *Gasterosteus aculeatus* correspond very well to probabilistic estimation using the social information. The link found between estimation and collective behavior should help to design experiments of collective behavior testing for the importance of estimation as a basic property of how brains work.

*Instituto Cajal, Consejo Superior de Investigaciones Científicas, Av. Doctor Arce, 37, 28002 Madrid, Spain. Department of Theoretical Physics and Instituto "Nicolás Cabrera" de Física de Materiales, Universidad Autónoma de Madrid, 28049 Madrid, Spain. E-mail: alfonso.perez.escudero@cajal.csic.es, gonzalo.polavieja@cajal.csic.es

Introduction

Animals need to make decisions without certainty in which option is best. This uncertainty is due to the ambiguity of sensory data but also to limited processing capabilities, and is an intrinsic and general property of the representation that animals can build about the world. A general way to make decisions in uncertain situations is to make probabilistic estimations [1, 2]. There is evidence that animals use probabilistic estimations, for example in the early stages of sensory perception [3–11], sensory-motor transformations [12–14], learning [15–17] and behaviors in an ecological context such as strategies for food patch exploitation [18–20] and mate selection [21], among others [13, 17, 21, 22].

An additional source of information about the environment may come from the behavior of other animals (social information) [23–28]. This information can have different degrees of ambiguity. In particular cases, the behavior of conspecifics directly reveals environmental characteristics (for example, food encountered by another individual informs about the quality of a food patch). Cases in which social information correlates well with the environmental characteristic of interest have been very well studied [29–37]. But in most cases social information is ambiguous and potentially misleading [26, 38]. In spite of this ambiguity, there is evidence that in some cases such as predator avoidance [39, 40] and mate choice [41], animals use this kind of information.

Social animals have a continuous flow of information about the environment coming from the behaviours of other animals. It is therefore possible that social animals use it at all times, making probabilistic estimations to counteract its ambiguity. If this is the case, estimation of the environment using both non-social and social information might be a major determinant of the structure of animal collectives. In order to test this hypothesis, we have developed a Bayesian decision-making model that includes both personal and social information, that naturally weights them according to their reliability in order to get a better estimate of the environment. All members of the group can then use these improved estimations to make better decisions, and collective patterns of decisions then emerge from these individuals interacting through their perceptual systems.

We show that this model derives social rules that economically explain detailed experiments of decision-making in animal groups [42, 43]. This approach should complement the empirical approach used in the study of animal groups [42–47], finding which mathematical functions should correspond to each experimental problem and to propose experiments relating estimation and collective motion. The Bayesian structure of our model also builds a bridge between the field of collective behavior and other fields of animal behavior, such as optimal foraging theory [18–22] and others [21, 22]. Further, it explicitly includes in a natural way different cognitive abilities, making more direct contact with neurobiology and psychology [3–10, 17].

Results

Estimation model

We derived a model in which each individual decides from an estimation of which behavior is best to perform. These behaviors can be to go to one of several different places, to choose among some behaviors like forage, explore or run away, or any other set of options. For clarity, here we particularize to the case of choosing the best of two spatial locations, x and y (see *Text S1* for more than two options). ‘Best’ may correspond to the safest, the one with highest food density or most interesting for any other reasons. We assume that each decision maker uses in the estimation of the best location both non-social and social information. Non-social information may include sensory information about the environment (i.e. shelter properties, potential predators, food items), memory of previous experiences and internal states. Social information consists of the behaviors performed by other decision-makers. Each individual estimates the probability that each location, say y , is the best one, using its non-social information (C) and the behavior of the other individuals (B),

$$P(Y|C, B), \quad (1)$$

where Y stands for ‘ y is the best location’. $P(X|C, B) = 1 - P(Y|C, B)$, because there are only two locations to choose from. We can compute the

probability in Eq. 1 using Bayes' theorem,

$$P(Y|C, B) = \frac{P(B|Y, C)P(Y|C)}{P(B|X, C)P(X|C) + P(B|Y, C)P(Y|C)}. \quad (2)$$

By simply dividing numerator and denominator by the numerator we find an interesting structure,

$$P(Y|C, B) = \frac{1}{1 + aS}, \quad (3)$$

where

$$a = \frac{P(X|C)}{P(Y|C)} \quad (4)$$

and

$$S = \frac{P(B|X, C)}{P(B|Y, C)}. \quad (5)$$

Note that a does not contain any social information so it can be understood as the “non-social term” of the estimation. We can also understand S as the “social term” because it contains all the social information, although it also depends on the non-social information C . The non-social term a is the likelihood ratio for the two options given only the non-social information. This kind of likelihood ratio is the basis of Bayesian decision-making in the absence of social information [5, 11–14]. Eq. 3 now tells us that this well known term interacts with the social term S simply through multiplication.

We are seeking a model based on probabilistic estimation that can simultaneously give us insight into social decision-making and fit experimental data. For this reason we simplify the model by assuming that the focal individual does not make use of the correlations among the behaviour of others, but instead assumes their behaviours to be independent of each other. This is a strong hypothesis but allows us to derive simple explicit expressions with important insights. The section ‘Model including dependencies’ at the end of Results shows that this assumption gives a very good approximation to a more complete model that takes into account these correlations.

The assumption of independence translates in that the probability of a given set of behaviors is just the product of the probabilities of the individual behaviors. We apply it to the probabilities needed to compute S in Eq. 5, getting

$$P(B|Y, C) = Z \prod_{i=1}^N P(b_i|Y, C), \quad (6)$$

where B is the set of all the behaviors of the other N animals at the time the focal individual chooses, $B = \{b_i\}_{i=1}^N$, and b_i denotes the behavior of one of them, individual i . Z is a combinatorial term counting the number of possible decision sequences that lead to the set of behaviors B , that will cancel out in the next step. Substituting Eq. 6 and the corresponding expression for $P(B|X, C)$ into Eq. 5, we get

$$S = \prod_{i=1}^N \frac{P(b_i|X, C)}{P(b_i|Y, C)}. \quad (7)$$

Instead of an expression in terms of as many behaviors as individuals, it may be more useful to consider a discrete set of behavioral classes. For example, in our two-choice example, these behavioral classes may be ‘go to x ’ (denoted β_x), ‘go to y ’ (β_y) and ‘remain undecided’ (β_u). Frequently, these behavioral classes (or simply ‘behaviors’) will be directly related to the choices, so that each behavior will consist of choosing one option. For example, behaviors β_x and β_y are directly related to choices x and y , respectively. But there may be behaviors not related to any option as the case of indecision, β_u , or related to choices in an indirect way. These behaviors can still be informative because they may be more consistent with one of the options being better than the other (for example, indecision may increase when there is a predator, so the presence of undecided individuals may bias the decision against the place where the non-social information suggests the presence of a predator). Let us consider L different behavioral classes, $\{\beta_k\}_{k=1}^L$. We do not here consider individual differences for animals performing the same behavior (say, behavior β_1), so they have the same probabilities $P(\beta_1|X, C)$ and $P(\beta_1|Y, C)$. Thus, if for example the n_1 first individuals are performing behavior β_1 , we have that $\prod_{i=1}^{n_1} \frac{P(b_i|X, C)}{P(b_i|Y, C)} = \left(\frac{P(\beta_1|X, C)}{P(\beta_1|Y, C)} \right)^{n_1}$. We can then write Eq. 7 as

$$S = \prod_{k=1}^L s_k^{n_k}, \quad (8)$$

where n_k is the number of individuals performing behavior β_k , and

$$s_k = \frac{P(\beta_k|X, C)}{P(\beta_k|Y, C)}. \quad (9)$$

The term s_k is the probability that an individual performs behavior β_k when x is the best option, over the probability that it performs the same behavior when y is the best choice. The higher s_k the more reliably behavior β_k indicates that x is better than y , so we can understand s_k as the reliability parameter of behavior β_k . If $s_k = \infty$, observing behavior β_k indicates with complete certainty that x is the best option, while for $s_k = 1$ behavior β_k gives no information. For $s_k < 1$, observing behavior β_k favors y as the best option, and more so the closer it is to 0. Note that $P(\beta_k|X, C)$ and $P(\beta_k|Y, C)$ are not the actual probabilities of performing behavior β_k , but estimates of these probabilities that the deciding animal uses to assess the reliability of the other decision-makers. These estimates may be ‘hard-wired’ as a result of evolutionary adaptation, but may also be subject to change due to learning.

To summarize, using Eqs. 3 and 8, the probability that y is the best choice, given both social and non-social information is

$$P(Y|C, B) = \left(1 + a \prod_{k=1}^L s_k^{n_k}\right)^{-1}, \quad (10)$$

with a in Eq. 4 and s_k in Eq. 9.

Decision rule: Probability matching

We have so far only considered the perceptual stage of decision-making, in which the deciding individual estimates the probability that each behavior is the best one. Now it must decide according to this estimation. A simple decision rule would be to go to y when $P(Y|C, B)$ is above a certain threshold. This rule maximizes the amount of correct choices when the probabilities do not change [48], but is not consistent with the experimental data considered in this paper. Applying this deterministic rule strictly, without any noise sources, one would obtain that all individuals behave exactly in the same way when facing the same stimuli, but in the experiments considered here this is not the case. Instead, we used a different decision rule called probability matching, that has been experimentally observed in many species, from insects to humans [49–55]. According to this rule an individual chooses each option with a probability that is equal to the probability that it is the best choice. Therefore, in our case the probability of going to y (P_y), is the same

as the estimated probability that y is the best location ($P(Y|C, B)$), so

$$P_y = P(Y|C, B). \quad (11)$$

Probability matching does not maximize the amount of right choices if we assume that the probabilities stay always the same, but in many circumstances it can be the optimal behavior, such as when there is competition for resources [56, 57], when the estimated probabilities are expected to change due to learning [53, 55], or for other reasons [53, 58].

Finally, using Eqs. 10 and 11 we have that the probability that the deciding individual goes to y is

$$P_y = \left(1 + a \prod_{k=1}^L s_k^{n_k}\right)^{-1}. \quad (12)$$

The assumption of probability matching has the advantage that the final expression for the decision in Eq. 12 is identical to the one given by Bayesian estimation in Eq. 10, with no extra parameters. Alternative decision rules could be noisy versions of the threshold rule, but at the price of adding at least one extra parameter to describe the noise. Also, decision rules might not depend on estimation alone, but also on other factors or constraints. These more complicated rules fall beyond the scope of this paper.

In the following sections, we particularize Eq. 12 to different experimental settings to test its results against existing rich experimental data sets that have previously been fitted to different mathematical expressions [42, 43].

Symmetric set-up

We first considered the simple case of two identical equidistant sites, x and y , Fig. 1A. For a set-up made symmetric by experimental design there is no true best option. But deciding individuals must act, like for any other case, using only their incomplete sensory data to make the best possible decision. Even when non-social sensory data indicates no relevant difference between the two sites, the social information can bias the estimation of the best option to one of the two sites.

Using Eq. 12 and that the three possible behaviors are ‘go to x ’ (β_x), ‘go to y ’ (β_y) and ‘remain undecided’ (β_u), we obtain

$$P_y = \left(1 + a s_x^{n_x} s_y^{n_y} s_u^{N-n_x-n_y}\right)^{-1}, \quad (13)$$

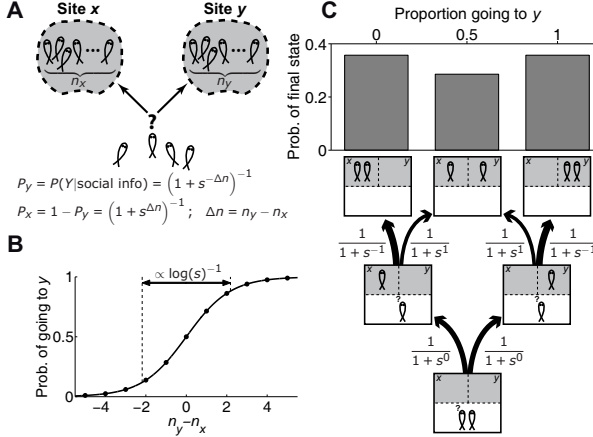


Figure 1. Model with individuals estimating which of two identical places is best. (A) Schematic diagram of individuals choosing between two identical locations x and y when there are already n_x (n_y) individuals at x (y). (B) Probability of going to y as a function of the difference between the number of individuals at y and x , Eq. 17. (C) Sequential application of the behavioural rule in Eq. 17 with $s = 2.5$, for the simple case of a group of two individuals (bottom). The width of the arrows is proportional to the probability of each transition. The 3 possible final configurations, with different proportion of individuals going to y (0, 0.5 and 1), have different probabilities of taking place, with both fish together at x or y being more probable than a group split (top).

where n_x and n_y are the number of individuals that have already chosen x and y , respectively, and $N + 1$ is the size of the group containing our focal individual and other N animals. As the setup is symmetric, the sensory information available to the deciding individual is the same for both options so $P(X|C) = P(Y|C)$ and then $a = 1$ according to Eq. 4. Also, since indecision is not related to any particular choice, symmetry imposes $P(\beta_u|X, C) = P(\beta_u|Y, C)$, so indecision is not informative, $s_u = 1$ (Eq. 9). For the other two behaviors, going to x (β_x) and going to y (β_y), Eq. 9 gives

$$\begin{aligned} s_x &= \frac{P(\beta_x|X, C)}{P(\beta_x|Y, C)} \\ s_y &= \frac{P(\beta_y|X, C)}{P(\beta_y|Y, C)}. \end{aligned} \quad (14)$$

$P(\beta_x|X, C)$ and $P(\beta_y|Y, C)$ are the estimated probabilities of making the right choice, that is, going to x when x is the best option, or going to y when y is the best option. Since in this case the sensory

information is identical for both options, the probability of making the correct choice must be the same for both options, $P(\beta_x|X, C) = P(\beta_y|Y, C)$. An analogous argument holds for the incorrect choices, $P(\beta_x|Y, C) = P(\beta_y|X, C)$, giving

$$s_x = 1/s_y. \quad (15)$$

In cases in which $s_x = 1/s_y$, we find it convenient to express reliability more generally as

$$s \equiv s_x = 1/s_y, \quad (16)$$

which is the ratio of the probability of making the correct choice and the probability of making a mistake, for both behaviors. Using this definition and given that $a = a_u = 1$, Eq. 13 reduces to

$$P_y = (1 + s^{-\Delta n})^{-1}, \quad (17)$$

with the variable $\Delta n \equiv n_y - n_x$. Eq. 17 describes a sigmoidal function that is steeper the higher the value of s (Fig. 1B). Therefore, for very reliable behaviors (high s , meaning individuals that are much more likely to make correct choices than erroneous ones), P_y grows fast with Δn and the deciding individual then goes to y with high probability when taking into account the behaviors of only very few individuals.

The behavior of the group is obtained by applying the decision rule in Eq. 17 sequentially to each individual (see *Methods*). After each behavioural choice, we update the number of individuals at x and y , using the new n_x and n_y for the next deciding individual (Fig. 1C, bottom). Repeating this procedure for all the individuals in the group, we can compute the probability for each possible final outcome of the experiment (Fig. 1C, top).

The relevance of the symmetric case is that the model has a single parameter and a single variable, enabling a powerful comparison against experimental data. We tested the model using an existing rich data set of collective decisions in three-spined sticklebacks [42], a shoaling fish species. This data set was obtained using a group of N_{tot} fish choosing between two identical refugia, one on their left and another one on their right (Fig. 2A), equivalent to locations x and y in the model (Fig. 1A). At the start of the experiment, m_x (m_y) replica fish made of resin were moved along lines on the left (right)

towards the refugia (Fig. 2A). The experimental results consisted on the statistics of collective decisions between the two refugia for 19 different cases using different group sizes $N_{\text{tot}} = 2, 4$ or 8 and different numbers of replicas going left and right, $m_x : m_y = \{1:1, 2:2, 0:1, 1:2, 0:2, 1:3, 0:3\}$ (Fig. 2B, blue histograms). To compare against these experimental data, we calculated the probability of finding a collective pattern applying the individual behavioural rule in Eq. 17 iteratively over each fish for the 19 experimental settings. We found a good fit of the model to the experimental data using for the 19 graphs the same value $s = 2.2$ (Fig. 2B, red line). The model is robust, with good fits in the interval $s = 2-4$ (Fig. 3, red line).

Despite the simplicity of the behavioral rule in Eq. 17, it reproduces the experimental results, including the dependence on the total number of fish N_{tot} , even though the rule is independent of this parameter, except for determining the range of possible values of Δn . The dependence of the final distributions on N_{tot} emerges from the application of the rule to the N_{tot} individuals in the group, as is illustrated in Fig. 4. Each small box represents a state of the system in which $n_x : n_y$ fish have already decided to go to x and y , respectively. The lines connecting each box with another two boxes on top represent the decision made by the next deciding individual, that takes the system to the next state. The width of the lines is proportional to the probability of the decision. As more individuals decide, the central states become less likely simply because they accumulate more unlikely decisions. Therefore, the U-shape or J-shape becomes more pronounced for larger groups, even though the individual decision rule in Eq. 17 is independent of the total number of individuals N_{tot} .

Group decision-making in three-spined sticklebacks shows a single type of distribution in which probability is minimum at the center and increases monotonically towards the edges, denoted here as U-shaped distribution (or J-shaped when there is a bias to one of the two options). However, the model in Eq. 17 also gives two other types of distributions, Fig. 5A. For non-social behavior ($s = 1$) the histogram is bell-shaped due to combinatorial effects. However, a bell-shape is also compatible with social animals for a certain range of s and group size (white region on the bottom-left of

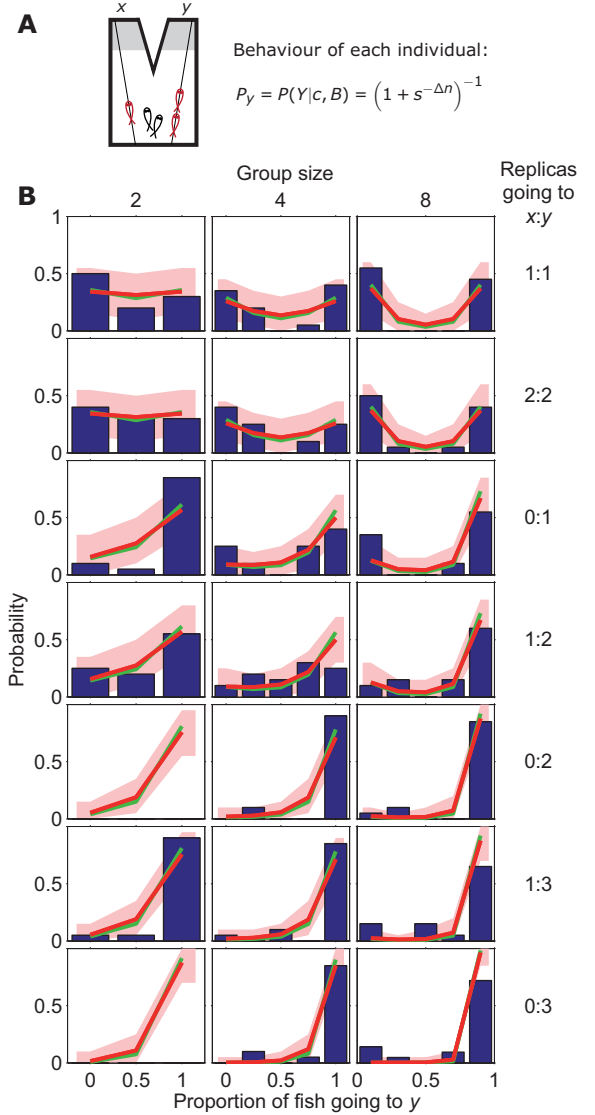


Figure 2. Comparison between model and stickleback choices in symmetric set-up. (A) Schematic diagram of symmetric set-up with a group of sticklebacks (in black) choosing between two identical refugia and with different numbers of replica fish (in red) going to x and y . (B) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histogram) and results from the model in Eq. 17 in the main text (red line using reliability parameter $s = 2.2$; red region: 95% confidence interval; green line with $s = 2.5$). Different graphs correspond to different stickleback group sizes and different number of replicas going to x and y .

Fig. 5A). For higher values of s , the histograms are

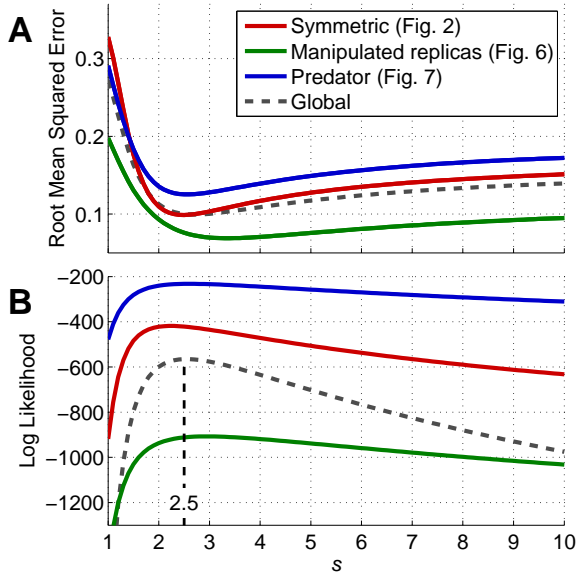


Figure 3. Goodness of fit for different values of the reliability (s). **Red:** Symmetric case (plots in Fig. 2). **Green:** Case with different replicas at each side (plots in Fig. 6). The ratios s_r/s_R are re-optimized for each value of s . **Blue:** Asymmetric set-up with predator on one side (plots in Fig. 7; Parameter a is re-optimized for each value of s). (A) Root mean squared error between the data and the probabilities predicted by the model. Grey dashed line shows the mean RMSE for the three cases. The absolute values for each case depend on the shape of the data and are not comparable, only the trends and the position of the minima should be compared. (B) Logarithm of the probability that the data come from the model. The height of each curve depends on the number of data for each experiment, only the trend and the position of the maxima should be compared. Grey dashed line shows the sum of the three coloured lines, but shifted by 1000 so that it fits on the scale. The peak of this global probability indicates the value of s that best fits the three datasets ($s = 2.5$).

M-shaped, with two maxima located between the center and the sides (region coloured in black and blue in Fig. 5A). However, the M shape becomes clear only with enough number of bins because the drop in probability near the edge or at the center of the distribution disappears when binning is too coarse, producing a bell-shaped or U-shaped histogram, Fig. 5B. This is an important practical issue, because the amount of data that can be collected rarely allows for more than 5 bins. The colorscale in Fig. 5A reflects the number of bins needed to observe the M shape (black has been reserved for

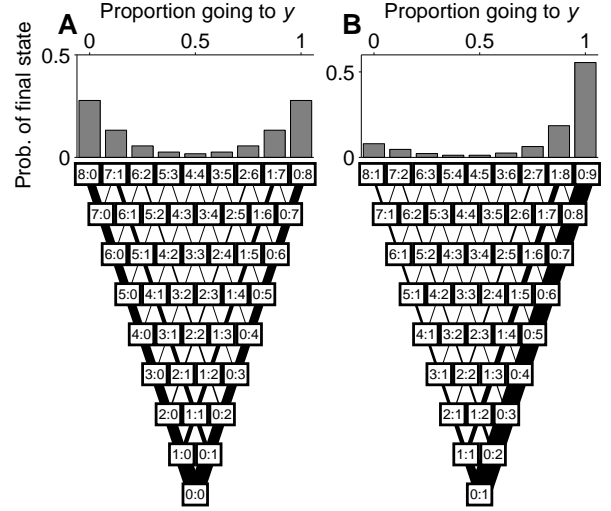


Figure 4. Illustration of the decision-making process in the model. **Bottom:** Decision-making process according to Eq. 17 (with $s = 2.5$). Time runs from bottom to top. Each box represents a state with a given number of fish having already decided x or y ($n_x : n_y$). Each state can lead to another two states in the following time step, depending on whether the focal fish decides to go to x or y . The width of the lines connecting states is proportional to the probability of that transition (equal to the probability of the prior state times the probability of the focal fish making the decision that leads to the later one). **Top:** Probability of each state after 8 fish have made their decisions. (A) Case with no replicas, in which the final outcome is U-shaped. (B) Case with one replica going to y (so initial state is already 0:1), in which the final outcome is J-shaped.

exactly 5 bins). For high values of s , the histograms are U-shaped (white region on the top of Fig. 5A). Also, all the M-region above the black zone becomes of type U when the binning is too coarse.

An interesting prediction of our model is that, for a given number of bins, the shape of the distribution of choices changes with the number of decided individuals, and the dynamics of this change depends on s . For high values of s , the probability is U-shaped from the beginning and becomes steeper as more individuals decide (as is the case for the stickleback dataset), Fig. 5C. For lower values of s , we observe M-shaped distributions for the first individuals and then U-shaped ones when more individuals decide, Fig. 5D. For even lower values of s , we observe bell-shaped distributions for the first individuals, then M-shaped and finally U-shaped, Fig. 5E,F.

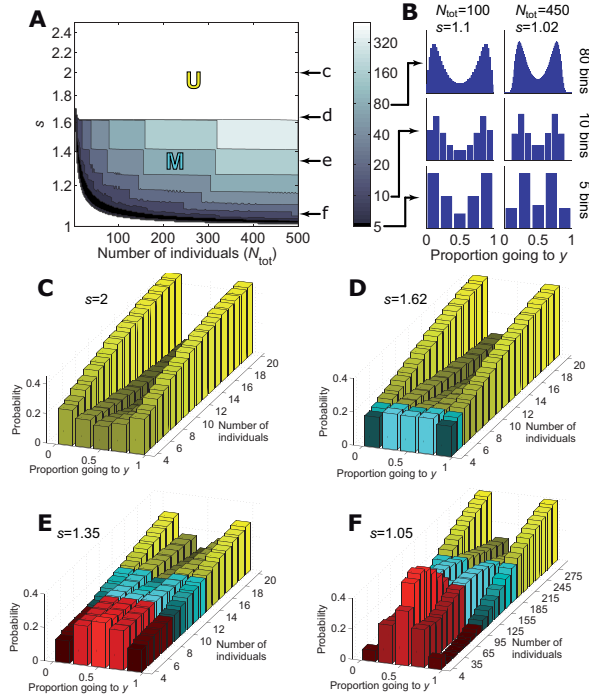


Figure 5. Types of distributions and dynamics for different values of reliability parameter s and group size. (A) Shape of histogram of final configurations as a function of s and the group size. Bell-shaped: white region on the bottom-left. M-shaped: region coloured in black and blue. As the observation of the M shape depends on the number of bins, the colorscale reflects the number of bins needed to observe the M shape (black has been reserved for exactly 5 bins). U-shape: white region on the top. Also, all the M-region above the black zone becomes U when the binning is too coarse. There is also a small region below the black zone where the M shape becomes a bell shape when the binning is too coarse. (B) Dependence of the apparent shape on the number of bins: Top, 80 bins. Middle, 10 bins. Bottom, 5 bins. On the left, a probability that seems U-shaped for 5 bins, but is M shaped for a higher number of bins. On the right, a probability that stays M-shaped for any number of bins. (C-F) Dynamics of the probability as the number of individuals increases for (C) $s = 2$, (D) $s = 1.62$, (E) $s = 1.35$ and (F) $s = 1.05$.

Symmetric set-up with modified replicas of animals

An interesting modification of the experimental set-up consists in using replicas of the animals that we can modify to potentially alter their reliability estimated by the animals. We considered the particular case, motivated by experiments in [43], of two types of modified replicas with different char-

acteristics (for example, fat or thin), Fig. 6A. We considered 7 behaviors: ‘animal goes to x ’ (β_{fx}), ‘animal goes to y ’ (β_{fy}), ‘most attractive replica goes to x ’ (β_{Rx}), ‘most attractive replica goes to y ’ (β_{Ry}), ‘least attractive replica goes to x ’ (β_{rx}), ‘least attractive replica goes to y ’ (β_{ry}), and ‘animal remains undecided’ (β_{fu}). The probability of going to y in Eq. 12 then reduces to

$$P_y = \left(1 + a s_{fx}^{n_{fx}} s_{fy}^{n_{fy}} s_{Rx}^{n_{Rx}} s_{Ry}^{n_{Ry}} s_{rx}^{n_{rx}} s_{ry}^{n_{ry}} s_{fu}^{N_f - n_{fx} - n_{fy}} \right)^{-1}, \quad (18)$$

where subindex ‘f’ refers to real fish and ‘R’ (‘r’) to replicas of the most (least) attractive type. As in the previous section, symmetry imposes that $a = 1$ and $s_{fu} = 1$. It also imposes the following relations between the reliability parameters, $s_f \equiv s_{fx} = 1/s_{fy}$, $s_R \equiv s_{Rx} = 1/s_{Ry}$, $s_r \equiv s_{rx} = 1/s_{ry}$. Therefore,

$$P_y = \left(1 + s_f^{-\Delta n_f} s_R^{-\Delta n_R} s_r^{-\Delta n_r} \right)^{-1}, \quad (19)$$

where $\Delta n_f \equiv n_{fy} - n_{fx}$, $\Delta n_R \equiv n_{Ry} - n_{Rx}$ and $\Delta n_r \equiv n_{ry} - n_{rx}$. In the particular case of only two different replicas, one going to x and the other to y and for notational simplicity taking the convention that the most (least) attractive replica goes to y (x), we have $\Delta n_R = 1$ and $\Delta n_r = -1$. Therefore,

$$P_y = \left(1 + \frac{s_r}{s_R} s_f^{-\Delta n_f} \right)^{-1}. \quad (20)$$

Note that the probability in Eq. 20 does not depend on s_r and s_R separately, but only on their ratio. Therefore, in this case the model uses only two parameters (s_f and s_r/s_R). We compared the model with the stickleback data set from [43], Fig. 6. The data in Fig. 6B has a different type of replica pair in each row, so in principle we would fit a different ratio s_r/s_R for each row. But note that the first three rows correspond to experiments with the same three replicas (large, medium and small), combined in different pairs. The same can be said for the second and third threesomes of rows. Therefore, there are only two free parameters for each three rows. On the other hand, s_f should have the same value for all cases. The model again reproduces the experimental results reported in reference [43], obtaining the best fit for $s_f = 2.9$ (Fig. 6B). The result is robust,

with good fits for $s_f = 2-4$ (Fig. 3, green line) in accord with the value obtained for the case shown in Fig. 2B.

Asymmetric set-up

We finally considered the case in which sites x and y are different and the three behaviors are ‘go to x ’ (β_x), ‘go to y ’ (β_y) and ‘remain undecided’ (β_u). Eq. 12 reduces to

$$P_y = (1 + a s_x^{n_x} s_y^{n_y} s_u^{N-n_x-n_y})^{-1}. \quad (21)$$

The term $a = P(X|C)/P(Y|C)$ represents the non-social information and in general $a \neq 1$ because the set-up is asymmetric by design. This asymmetry might also affect how a deciding animal takes into account the behaviours of other animals depending on which side they chose, making in general $s_x \neq 1/s_y$. Also, indecision might be informative. For example, if non-social information indicates the possible presence of a predator at y , the indecision of other animals might confirm this to the deciding individual, further biasing the decision towards x . Therefore, we may have $s_u \neq 1$. But it may also be the case that the set-up’s asymmetry does not affect the social terms, so we also tested a simpler model in which $s \equiv s_x = 1/s_y$ and $s_u = 1$, giving

$$P_y = (1 + a s^{-\Delta n})^{-1}. \quad (22)$$

The stickleback dataset reported in reference [42] is ideally suited to test the asymmetric model for the experiments that were performed with a replica predator at the right arm (Fig. 7A). The model in Eq. 22 fits best the data with $s = 2.6$ (Fig. 7B) and it is robust with a good fit in $s = 2-4$ (Fig. 3, blue line). The more complex model in Eq. 21 gives fits very similar to those of simpler model. Specifically, parameter s_u was rejected by the Bayes Information Criterion [59, 60], suggesting that fish do not rely on undecided individuals. The fact that fish rely differently on other fish depending on the option they have taken could not be ruled out by the Bayes Information Criterion, but in any case the impact of this difference on the data is small.

In the experiments in Fig. 2 and Fig. 7, we have assumed that the replicas are perceived by fish as real animals. However, it is reasonable to think that fish might perceive the difference, and rely differently on replicas and real fish. To test this, we

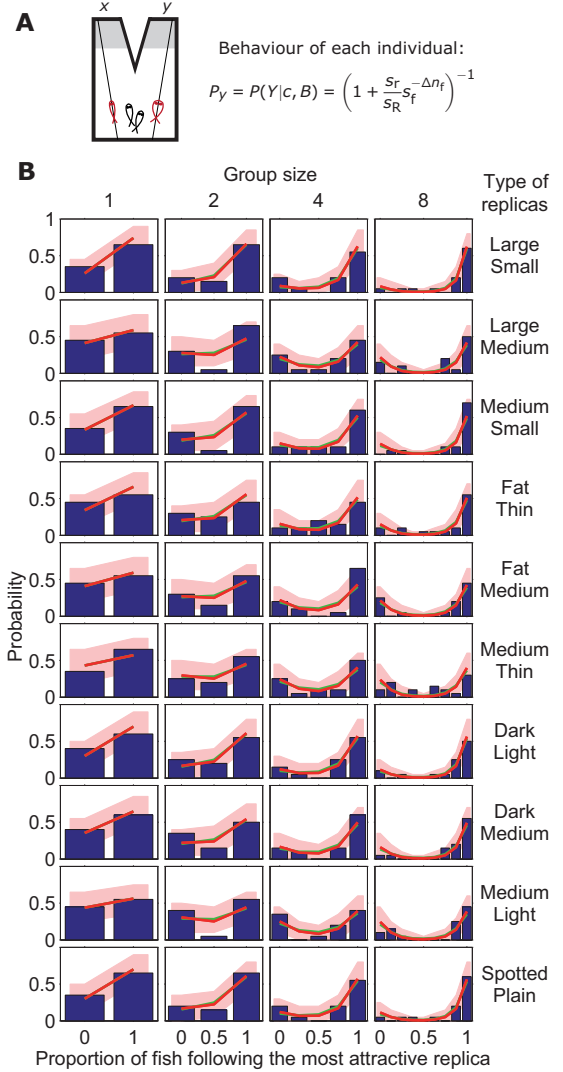


Figure 6. Comparison between model and stickleback choices with two differently modified replicas. (A) Schematic diagram of symmetric set-up with a group of sticklebacks (in black) choosing between two identical refugia and with one replica fish going to x and a different one (in size, shape or pattern) going to y (in red). (B) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [43] (blue histogram) and results from model in Eq. 20 in the main text (red line using reliability parameter $s_f = 2.9$ and $s_r/s_R = 0.35, 0.7, 0.5, 0.52, 0.69, 0.75, 0.43, 0.55, 0.78, 0.43$, for each row from top to bottom; red region: 95% confidence interval; green line with $s_f = 2.5$ and same ratios s_r/s_R as for red line). Different graphs correspond to different stickleback group sizes and different types of replicas going to x and y .

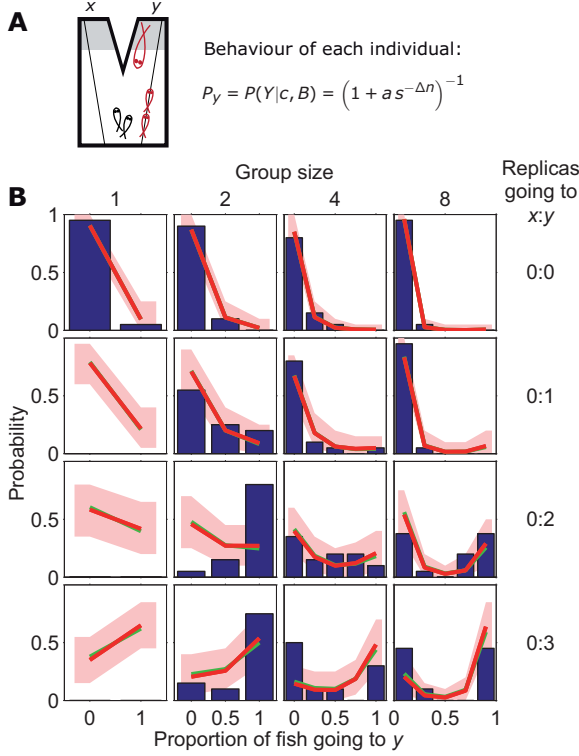


Figure 7. Comparison between model and stickleback choices in asymmetric set-up. (A) Schematic diagram of asymmetric set-up (predator at y , large fish depicted in red) with a group of sticklebacks (in black) choosing between two refugia, and replica fish (small fish depicted in red) going to y . (B) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histogram) and results from model in Eq. 22 in the main text (red line using $s = 2.6$, $a = 9.5$; red region: 95% confidence interval. Green line using $s = 2.5$ and same a as for red line). Different graphs correspond to different stickleback group sizes and different number of replicas going to y .

considered different behaviors for fish and replicas, such as ‘fish goes to x ’ and ‘replica goes to x ’. Making that distinction, we get that Eq. 12 reduces to

$$P_y = \left(1 + a s_{fx}^{n_{fx}} s_{fy}^{n_{fy}} s_{rx}^{n_{rx}} s_{ry}^{n_{ry}} s_{fu}^{N_t - n_{fx} - n_{fy}} \right)^{-1}. \quad (23)$$

The Bayes Information Criterion rejects only parameter s_{fu} . However, the addition of the new parameters that distinguish replica from real fish give very small improvements in the fits compared to results of the simpler models in Eq. 17 and Eq. 22 (see Fig. S1 and S3), suggesting that fish follow replicas

as much as they follow real fish.

Model including dependencies

In this section we will remove the hypothesis of independence among the behaviors of the other individuals (Eq. 6). We now consider that the focal individual not only takes into account the behaviors of the other animals at the time of decision but the specific sequence of decisions that has taken place before, $\{b_i\}_{i=1}^{K-1}$, being $K - 1$ the number of individuals that have decided before the focal one. For example, the sequence $\{x, y\}$ may give different information to the focal individual than the sequence $\{y, x\}$. This is illustrated in Fig. 8A, where there are two possible paths leading to states labeled as 1:1, but these two states are in different branches of the tree (in contrast with Fig. 4, in which these two states were collapsed in a single one).

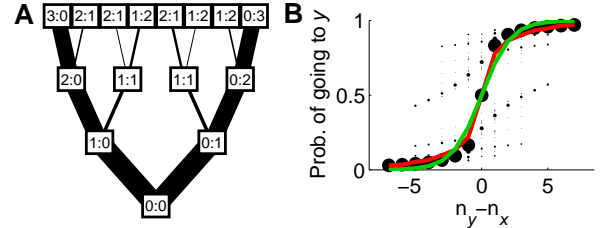


Figure 8. Model taking into account dependencies. (A) Decision-making process according to the model with dependencies, Eq. 25-33. Time runs from bottom to top. Each box represents one state, and each edge represents one option of the deciding individual, that either goes to x or to y . Edge width is proportional to the probability of the decision. (B) Probability of choosing y as a function of the difference of the number of individuals that have already chosen each option ($\Delta n = n_y - n_x$), for $a'_X = 5$. In the new model the probability does not depend any more on Δn alone, so states with the same Δn have different values for the probability (black dots). The area of the dots is proportional to the probability of observing each state. Red line shows the expected value of the probability for each value of Δn . The green line shows the probability for the model that neglects dependencies (Eq. 17), $(1 + s^{-\Delta n})^{-1}$ for $s = 2.5$.

To calculate the probability of the observed sequence of behaviors provided that Y is the correct choice,

$$P(\{b_i\}_{i=1}^{K-1} | Y, C_K), \quad (24)$$

one can apply $P(A, B) = P(A|B)P(B)$ repeatedly

to obtain

$$P(\{b_i\}_{i=1}^{K-1}|Y, C_K) = \prod_{k=1}^{K-1} P(b_k|Y, C_K, \{b_i\}_{i=1}^{k-1}). \quad (25)$$

This expression substitutes the assumption of independence in Eq. 6. Each of the terms in the product is simply the probability that the k -th individual makes its decision, given the previous decisions, and also given that y is the correct choice. This result was expected since if we look at the tree in Fig. 8A we see that the probability of reaching a given state is simply the product of the probabilities of choosing the adequate branches in each step.

So the problem reduces to computing the individual decision probabilities $P(b_k|Y, C_K, \{b_i\}_{i=1}^{k-1})$. We assume in the following that these probabilities are calculated by the focal individual by assuming that all animals use the same rules to make a decision. The rule for the focal individual is, as in previous sections,

$$P_{yK} = P(Y|\{b_i\}_{i=1}^{K-1}, C_K) = \frac{1}{1 + a_K S_K}, \quad (26)$$

where the non-social and social terms are

$$a_K = \frac{P(X|C_K)}{P(Y|C_K)}, \quad (27)$$

and

$$S_K = \frac{P(\{b_i\}_{i=1}^{K-1}|X, C_K)}{P(\{b_i\}_{i=1}^{K-1}|Y, C_K)}, \quad (28)$$

respectively, and where we have added subscript K to S , a and C to reflect that they apply to the focal individual, that makes its decision in the K -th place.

The assumption that all animals apply the same rules translates into the following. To apply an equation like Eq. 26 but on a different individual (say, individual k) it is necessary to know the non-social information C_k . Remember that all these computations are made from the point of view of the focal individual, and obviously the focal individual does not have access to the non-social information of the other individuals. It may seem reasonable for the focal animal to assume that all the other individuals have the same non-social information (C_K), but this would result in no social behavior at all (if the other individuals have the same non-social information, their behaviors will not give any extra information). Instead, one can assume that the other

individuals may have a different non-social information, C' . Furthermore, this non-social information depends on which is the best choice, because if for example x is the best choice the other individuals have some probability of detecting it, and therefore their non-social information will be on average biased towards x . We approximate this average bias by assuming that, if x (y) is the best choice, all the other individuals will have non-social information C'_X (C'_Y) that will bias the decision towards x (y). It is therefore the same to assume that x (y) is the best option as to assume that all the other individuals have non-social information C'_X (C'_Y). Therefore, for the probabilities of individual behaviors in Eq. 25, we have that

$$P(b_k|Y, C_K, \{b_i\}_{i=1}^{k-1}) = P(b_k|C'_Y, C_K, \{b_i\}_{i=1}^{k-1}), \quad (29)$$

where now C'_Y applies to the k -th individual, so we can compute this probability simply by applying Eq. 26 to the k -th individual,

$$P_{y_k, Y} = \frac{1}{1 + a'_Y S_k}, \quad (30)$$

where

$$a'_Y = \frac{P(X|C'_Y)}{P(Y|C'_Y)}. \quad (31)$$

Then, if we denote $P_{b_k, Y} \equiv P(b_k|C'_Y, C_K, \{b_i\}_{i=1}^{k-1})$, we have that

$$\begin{cases} P_{b_k, Y} = P_{y_k, Y} & \text{if } b_k = y \\ P_{b_k, Y} = 1 - P_{y_k, Y} & \text{if } b_k = x. \end{cases} \quad (32)$$

These are the individual probabilities needed in Eq. 25, that takes into account the correlations among the other individuals. So we can already calculate S_k using Eq. 28,

$$S_k = \frac{\prod_{i=1}^{k-1} P_{b_i, X}}{\prod_{i=1}^{k-1} P_{b_i, Y}}, \quad (33)$$

Eqs. 30 and 33 have a recursive relation, because we need the probabilities up to step $k-1$ to compute S_k , and then we need S_k to compute the probabilities in step k . At the beginning no individual has made any choices, so we start with $S_1 = 1$ and work recursively from there until we obtain the probabilities for individual $K-1$, that allow to compute S_K . Then, we can already use Eq. 26 to compute the decision probability of the focal individual, this

time using its actual non-social term a_K (which is 1 for the symmetric cases, and fitted to the data in the non-symmetric case).

The equations above constitute the model taking into account dependencies. The new parameters of this model are a'_X and a'_Y , which substitute s_x and s_y in the previous models, so the number of parameters is exactly the same. In the symmetrical case we must have that $a'_X = 1/a'_Y$, so the model has a single parameter. For the non-symmetrical case these parameters may be independent of each other, but we find good results even assuming that they are not, as was the case for the simplified model. So for simplicity we always assume that

$$a'_X = 1/a'_Y. \quad (34)$$

For the case with different replicas at each side, each of them has a different value of a'_X , thus making one replica more attractive than the other.

The new model also matches very well with the experimental data discussed in this paper. Results for the case of two different replicas are shown in Fig. 9, for the symmetric case in Fig. S4 and for the case with predator in Fig. S5. Fits are robust, and all cases are well explained by the model with the same value of $a'_X = 5$, Fig. S6. See Figs. S1-S3 for a comparison of all models.

We now ask how different is the model including dependencies from the model that neglects them. To compare the two models, we plot the probability of going to y as a function of $\Delta n = n_y - n_x$ for the new model, as we did in Fig. 1B for the old one. The inclusion of dependencies has the consequence that the probability of going to y does not depend only on Δn , since now different states with the same Δn may have different probabilities. Therefore, when we plot the probability of going to y as a function of Δn we obtain different values of the probability for each value of Δn . This is shown by the black dots in Fig. 8B, where the size of the dots is proportional to the probability of observing each state when starting from 0:0. The red line shows the average probability for each Δn , taking into account the probability of each state. Both the dots and this line correspond to $a'_X = 5$, which is the one that fits best the data. The green line corresponds to the probability for the simplest model neglecting dependencies, with the value that best fits to the data ($s = 2.5$). This line is close

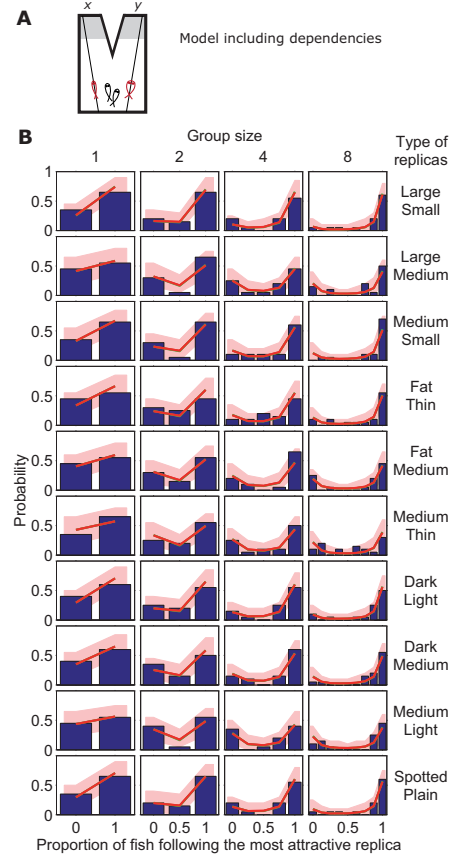


Figure 9. Comparison between model including dependencies and stickleback choices with two differently modified replicas. (A) Schematic diagram of symmetric set-up with a group of sticklebacks (in black) choosing between two identical refugia and with one replica fish going to x and a different one (in size, shape or pattern) going to y (in red). (B) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [43] (blue histogram) and results from model that takes dependencies into account (red line, with $a'_{X,\text{fish}} = 4.8$ and $a'_{X,\text{replicas}} = 21.4, 11.8, 0.6, 9.9, 4.8, 0.9, 13, 8, 0.7, 14.5, 0.9$, for each type of replica (large, medium, small, fat, etc.); red region: 95% confidence interval; green line with $a'_{X,\text{fish}} = 5$ and same $a'_{X,\text{replicas}}$ as for red line). Different graphs correspond to different stickleback group sizes and different types of replicas going to x and y .

to the mean probability for the new model and to the values with highest probability of occurrence, so the simple model is as a good approximation to the model with dependencies.

We find an interesting prediction of the new model: There are some states in which the most

likely option is to choose the option chosen by *fewer* individuals (for example, note in Fig. 8D that some points with $\Delta n < 0$ are above 0.5). This surprising result comes from the fact that, as more fish accumulate at one side, their choices become less and less informative (because it is very likely that they are simply following the others). If then one fish goes to the opposite side, its behavior is very informative, because it is contradicting its social information. This effect can be so strong that it may beat the effect of all the other individuals, resulting in a higher probability of following this last individual than all the individuals that decided before.

Discussion

We have shown that probabilistic estimation in the presence of uncertainty can explain collective animal decisions. This approach generated a new expression for each experimental manipulation, Eq. 17-22, and was naturally extended to test for more refined cognitive capacities, Eq. 23. The model was found to have a good correspondence with the data in three experimental settings (Figs. 2, 6 and 7), always giving a good fit with the social reliability parameter s in the interval 2-4. Indeed, all the data have a very good fit with $s = 2.5$ (Figs. 2, 6 and 7, green lines). According to Eq. 9, this value for s has the interpretation that, for the behaviors relevant for these experiments, the fish assume that their conspecifics make the right choice 2.5 times more often than the wrong choice.

For the data used in this paper, previous empirical fits used more parameters [42] (Figs. S1-S3, blue line), and added more complex behavioral rules when the basic model failed [43] (Fig. S2, blue line). Our approach thus gains in simplicity. It also finds an expression for each set-up with expressions for complex set-ups obtained with add-ons to those of simpler set-ups, making the model scalable and easier to understand in terms of simpler experiments. Also, taking the models as fits to experimental data, the bayesian information criterion finds our models to be better than those in [42] and [43] (see captions in Figs. S1-S3 for details).

Collective animal behavior has been subject to a particularly careful quantitative analysis. Previous studies have given descriptions led by the powerful idea that complex collective behaviors can emerge

from simple individual rules. In fact, some systems have been found empirically to obey rules that are mathematically similar or the same as some of the ones presented in this paper, further supporting the idea that probabilistic estimation might underlie collective decision rules in many species. For example, a function like the one in Eq. 17 has been used to describe the behavior of Pharaoh's ant [61], a function like Eq. 22 for mosquito fish [62], and a function like the one in the right-hand-side of Eq. 22 for meerkats [63]. But despite the importance of group decisions in animals, little is known about the origin of such simple individual rules. This paper argues that probabilistic estimation can be an underlying substrate for the rules explaining collective decisions, thus helping in their evolutionary explanation. Also, this connection between patterns in animal collectives and a cognitive process helps to explain the similarities that exist between decision-making processes at the level of the brain and at the level of animal collectives [64, 65].

Our model is naturally compatible with other theories that use a Bayesian formalism to study different aspects of behavior and neurobiology, thus contributing to a unified approach of information processing in animals. For example, it may be combined with the formalism of Bayesian foraging theory [18], through an expansion of the non-social reliability a . Related to this case, a very well studied example of use of social information is the one in which one individual can observe directly the food collected by another individual [29-33]. In this case the social information is as unambiguous as the non-social one, so in this case both types of information should have a similar mathematical form [29-33]. This is consistent with our model, that in this case will give a similar expression for a and S . Other kinds of social information (such as another individual's decision to leave a food patch or choices of females in mating [41]) would enter naturally in our reliability terms s_k . In discussing these and similar problems, it has been proposed that animals should use social information when their personal information is poor, and ignore it otherwise [25, 26, 41]. Our model provides a quantitative framework for this problem, predicting that social information is always used, only with different weights with respect to other sources of information. Bayesian estimation is also a prominent approach to study decisions

in neurobiology and psychology [3–17] and it would be of interest to explore the mechanisms and role played by the multiplicative relation between non-social and social terms.

Our approach also makes a number of predictions. For example, it derives the probability of choosing among M options (see Eq. S16 of the Text S1), that for the symmetric case reduces to

$$P_\mu = \left(1 + \sum_{m \neq \mu}^M s^{-(n_\mu - n_m)} \right)^{-1}, \quad (35)$$

predicted also to fit the data for cases with $M > 2$ options.

We also predict a quantitative link between estimation and collective behavior. The parameters a and s_k in our model are in fact not merely fitting parameters, but true experimental variables. Manipulations of a and s_k should allow to test that changes in collective behavior follow the predictions of the model. A counterintuitive prediction about the manipulation of s_k is that external factors unrelated to the social component can nevertheless modify it. For example, a fish that usually finds food in a given environment should interpret a sudden turn of one of his mates as an indication that it has found food, and therefore will follow it. In contrast, another fish that is not expected to find food in that environment will not interpret the sudden turn as indicative of food, and will not follow. Thus, the model predicts that the *a priori* probability of finding food (to which each fish can be trained in isolation) will modify its propensity to follow conspecifics. An alternative approach that would not need manipulation of the reliabilities s_k would consist in showing that the probability of copying a behavior increases with how reliably the behavior informs about the environment.

We can also extend the estimation model to use, instead of the location of animals, their predicted location. We would then find expressions like the ones in this paper but for the number or density of individuals estimated for a later time. Consider for example the case without non-social information, described in Eq. 17 for two options and in Eq. 35 for more options. We can rewrite these equations as $P_\mu = \Omega s^{n_\mu}$ with μ one of the options and Ω is the normalization, $\Omega = \sum_{m=1}^M s^{n_m}$, where M is the number of options. Then, we would have

$P(\vec{x}) = \Omega s^{\rho(\vec{x}; t + \Delta t)}$ for the continuous case using prediction. Future positions at times $t + \Delta t$ (where Δt does not need to be constant) in terms of variables at present time t would be given by $\vec{x} + \vec{v}\Delta t$ for animals moving at constant velocity \vec{v} . Consider then a simple case of an animal located at \vec{x} and estimating the future position of a compact group at \vec{x}_g and moving with velocity \vec{v}_g . The deciding animal would be predicted to move with a high probability in the direction $(\vec{x}_g(t) - \vec{x}(t)) + \Delta t \vec{v}_g(t)$. Estimation of future locations thus naturally predicts in this simple case a particular form of ‘attraction’ and ‘alignment’ forces of dynamical empirical models [46, 66] as attraction to future positions, but in the general also deviations from these simple rules.

Methods

Obtaining group behavior from the model of an individual

The estimation rules presented in this paper refer to a single individual. To simulate the behavior of a group, we use the following algorithm: The current individual decides between x and y . After the decision, we recompute the relevant parameters of the model and use the new values for the next deciding individual. The undecided individuals are only those that are waiting for their turn to decide. We tested an alternative implementation in which individuals may remain undecided or in which two individuals can decide simultaneously, obtaining no relevant differences.

For the case of the model including dependencies, the model always starts at state 0:0, with $S = 1$. Most experiments have initial conditions in which several replicas are already going to either side, and the fish have no information about the path followed to reach this state. In these cases, we average the probabilities of all the paths that might have possibly led to the initial state to compute the initial value of S .

Protocol S1 and *Protocol S2*, contain Matlab functions that run the models (extensions of the files must be changed from .txt to .m to make them operative). *Protocol S1* corresponds to the model without dependencies, and *Protocol S2* corresponds to the model with dependencies. These functions have been used to generate all the theoretical re-

sults presented in this paper.

Fits

We computed log likelihood as the logarithm of the probability that the histograms come from the model. We searched for the model parameters giving a higher value of log likelihood, corresponding to a better fit. This search was performed by optimizing each parameter separately (keeping the rest constant) and iterating through all parameters until convergence. In all cases convergence was rapidly achieved. We performed multiple searches for best fitting parameters starting from random initial conditions and always found convergence to the same values, suggesting there are no local maxima. Indeed, we observed that log-likelihood is smooth and with a single maximum in all the cases with 1 or 2 parameters (see Fig. 3 for an example).

Bayesian Information Criterion

For model comparison we used the Bayesian Information Criterion (BIC) [59, 60], which takes into account both goodness of fit and the number of parameters. According to this criterion, among several models that have been fitted to maximize log likelihood, one should select the one for which

$$BIC_i = L_i - \frac{1}{2}k_i \log(h) \quad (36)$$

is largest, where L_i is the logarithm of the probability that the data comes from the i -th model once its parameters have been optimized to maximize this probability, k_i is its number of parameters of the i -th model and h is the number of measurements (which in our case is the same for all models).

More intuitive than the direct BIC_i values in Eq. 36 are the BIC weights, defined as [60]

$$w_i = \frac{\exp(BIC_i)}{\sum_j \exp(BIC_j)}, \quad (37)$$

when we assume that all models are *a priori* equally likely. Roughly speaking, w_i can be interpreted as the probability that model i is the most correct one [60].

We used BIC to compare different versions of our model, and also to compare our model with those of references [42, 43] (see Figs. S1-S3). The models

of refs. [42, 43] were originally fitted by minimizing the mean squared error instead of by maximizing logprob. For this reason, they score very poorly in BIC with their reported parameters. For this reason, we re-optimized for maximum logprob all their model parameters (these parameters are, using the notation of refs. [42, 43], a , k , T , m and r , with r only applicable in the case of predator present). For the case of different replicas going to each side, parameter p_{bias} takes a different value for each row in the figure, adding up to 10 parameters. The model in ref. [43] is computationally expensive, so it is not feasible to re-optimize these many parameters. Therefore, we treated them as if they were independently measured: we fixed p_{bias} in each case so that the results of the trials with a single individual matched exactly the model's prediction (as reported in [43]). We also followed this procedure with the ratios s_r/s_R of our model without dependencies, and the pairs $a'_{X,\text{replicas}}$ in our model with dependencies. Then, we performed BIC taking into account neither these parameters (p_{bias} the ratios s_r/s_R and the pairs $a'_{X,\text{replicas}}$) nor the data from trials using single individuals.

Acknowledgments

We acknowledge useful comments by Sara Arganda, Larissa Conradt, Iain Couzin, Jacques Gautrais, David Sumpter, Guy Theraulaz, Julián Vicente Page and COLMOT 2010 participants.

References

1. Box G, Tiao G (1973) Bayesian inference in statistical analysis. New York: Addison-Wesley. URL <http://onlinelibrary.wiley.com/doi/10.1002/9781118033197.fmatter/summary>.
2. Jaynes ET, Bretthorst LG (2003) Probability Theory: The Logic of Science (Vol 1). Cambridge University Press.
3. Helmholtz H (1925) Physiological Optics, Vol. III: The perceptions of Vision. Rochester, NY, USA: Optical Society of America.
4. Mach E (1980) Contributions to the Analysis of the Sensations. Chicago, IL, USA: Open Court Publishing Co.
5. Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences* 27: 712–9.
6. Jacobs R (1999) Optimal integration of texture and motion cues to depth. *Vision Research* 39: 3621–3629.
7. Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research* 43: 2539–2558.
8. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415: 429–33.
9. Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A* 20: 1391.
10. Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Current Biology* 14: 257–262.
11. Gold JJ, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences* 5: 10–16.
12. Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. *Nature* 427: 244–247.
13. Kording KP, Wolpert DM (2006) Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences* 10: 319–26.
14. Gold JJ, Shadlen MN (2007) The neural basis of decision making. *Annual review of neuroscience* 30: 535–74.
15. Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences* 10: 294–300.
16. Kruschke JK (2006) Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological review* 113: 677–99.
17. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND (2011) How to Grow a Mind: Statistics, Structure, and Abstraction. *Science* 331: 1279–1285.
18. Oaten A (1977) Optimal foraging in patches: A case for stochasticity. *Theoretical Population Biology* 12: 263–285.

19. Biernaskie JM, Walker SC, Gegear RJ (2009) Bumblebees learn to forage like Bayesians. *The American naturalist* 174: 413–423.
20. Alonso J (1995) Patch use in cranes: a field test of optimal foraging predictions. *Animal Behaviour* 49: 1367–1379.
21. McNamara JM, Green RF, Olsson O (2006) Bayes theorem and its applications in animal behaviour. *Oikos* 112: 243–251.
22. Valone TJ (2006) Are animals capable of Bayesian updating? An empirical review. *Oikos* 112: 252–259.
23. Valone TJ, Templeton JJ (2002) Public information for the assessment of quality: a widespread social phenomenon. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 357: 1549–57.
24. Blanchet S, Clobert J, Danchin E (2010) The role of public information in ecology and conservation: an emphasis on inadvertent social information. *Annals of the New York Academy of Sciences* 1195: 149–68.
25. Dall SRX, Giraldeau LA, Olsson O, McNamara JM, Stephens DW (2005) Information and its use by animals in evolutionary ecology. *Trends in ecology & evolution* 20: 187–93.
26. Giraldeau LA, Valone TJ, Templeton JJ (2002) Potential disadvantages of using socially acquired information. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 357: 1559–66.
27. Wagner RH, Danchin E (2010) A taxonomy of biological information. *Oikos* 119: 203–209.
28. King AJ, Cowlshaw G (2007) When to use social information: the advantage of large group size in individual decision making. *Biology letters* 3: 137–9.
29. Valone TJ (1989) Group Foraging, Public Information, and Patch Estimation. *Oikos* 56: 357–363.
30. Templeton JJ, Giraldeau LA (1995) Patch assessment in foraging flocks of European starlings: Evidence for the use of public information. *Behavioral Ecology* 6: 65–72.
31. Templeton JJ, Giraldeau LA (1996) Vicarious sampling: The use of personal and public information by starlings foraging in a simple patchy environment. *Behavioral Ecology and Sociobiology* 38: 105–14.
32. Smith JW, Benkman CW, Coffey K (1999) The use and misuse of public information by foraging red crossbills. *Behavioral Ecology* 10: 54–62.
33. Clark C, Mangel M (1986) The evolutionary advantages of group foraging. *Theoretical Population Biology* 30: 45–75.
34. Doligez B, Danchin E, Clobert J (2002) Public information and breeding habitat selection in a wild bird population. *Science (New York, NY)* 297: 1168–70.
35. Boulinier T, Danchin E (1997) The use of conspecific reproductive success for breeding patch selection in terrestrial migratory species. *Evolutionary Ecology* 11: 505–517.
36. Coolen I, van Bergen Y, Day RL, Laland KN (2003) Species difference in adaptive use of public information in sticklebacks. *Proceedings Biological sciences / The Royal Society* 270: 2413–9.

37. van Bergen Y, Coolen I, Laland KN (2004) Nine-spined sticklebacks exploit the most reliable source when public and private information conflict. *Proceedings Biological sciences / The Royal Society* 271: 957–62.
38. Rieucan G, Giraldeau La (2009) Persuasive companions can be wrong: the use of misleading social information in nutmeg mannikins. *Behavioral Ecology* 20: 1217–1222.
39. Lima SL (1995) Collective detection of predatory attack by social foragers: fraught with ambiguity? *Animal Behaviour* 50: 1097–1108.
40. Proctor CJ, Broom M, Ruxton GD (2001) Modelling antipredator vigilance and flight response in group foragers when warning signals are ambiguous. *Journal of theoretical biology* 211: 409–17.
41. Nordell, Valone TJ (1998) Mate choice copying as public information. *Ecology Letters* 1: 74–76.
42. Ward AJW, Sumpter DJT, Couzin ID, Hart PJB, Krause J (2008) Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Sciences of the United States of America* 105: 6948–53.
43. Sumpter DJT, Krause J, James R, Couzin ID, Ward AJW (2008) Consensus decision making by fish. *Current Biology* 18: 1773–1777.
44. Couzin ID, Krause J (2003) Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior* 32: 1–75.
45. Sumpter DJT (2006) The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 361: 5–22.
46. Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. *Nature* 433: 513–516.
47. Katz Y, Tunstrom K, Ioannou CC, Huepe C, Couzin ID (2011) Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences* : 1–6.
48. Neyman J, Pearson E (1933) On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character* 231: 289.
49. Herrnstein R (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior* 4: 267.
50. Behrend ER, Bitterman ME (1961) Probability-Matching in the Fish. *The American Journal of Psychology* 74.
51. Greggers U, Menzel R (1993) Memory dynamics and foraging strategies of honeybees. *Behavioral Ecology and Sociobiology* 32: 17–29.
52. Kirk KL, Bitterman ME (1965) Probability-Learning by the Turtle. *Science* 148: 1484–1485.
53. Vulkan N (2000) An Economist’s Perspective on Probability Matching. *Journal of Economic Surveys* 14: 101–118.
54. Wozny DR, Beierholm UR, Shams L (2010) Probability matching as a computational strategy used in perception. *PLoS Computational Biology* 6: 7.

55. Staddon J (1983) *Adaptive Behavior and Learning*. Cambridge: Cambridge University Press. URL <http://dukespace.lib.duke.edu/dspace/handle/10161/2878>.
56. Fretwell S, Lucas H (1969) On territorial behavior and other factors influencing habitat distribution in birds. *Acta biotheoretica* 19: 16–36.
57. Houston A, McNamara J (1987) Switching between resources and the ideal free distribution. *Animal behaviour* 35: 301–302.
58. Gaissmaier W, Schooler LJ (2008) The smart potential behind probability matching. *Cognition* 109: 416–22.
59. Schwarz G (1978) Estimating the dimension of a model. *The annals of statistics* 6: 461–464.
60. Link W (2006) Model weights and the foundations of multimodel inference. *Ecology* 87: 2626–2635.
61. Jeanson R, Ratnieks FLW, Deneubourg JL (2003) Pheromone trail decay rates on different substrates in the Pharaoh’s ant, *Monomorium pharaonis*. *Physiological Entomology* 28: 192–198.
62. Ward AJW, Herbert-Read JE, Sumpter DJT, Krause J (2011) Fast and accurate decisions through collective vigilance in fish shoals. *PNAS* 108: 6–9.
63. Bousquet CAH, Sumpter DJT, Manser MB (2010) Moving calls: a vocal mechanism underlying quorum decisions in cohesive groups. *Proceedings of the Royal Society B Biological Sciences* in press.
64. Marshall JaR, Bogacz R, Dornhaus A, Planqué R, Kovacs T, et al. (2009) On optimal decision-making in brains and social insect colonies. *Journal of the Royal Society, Interface / the Royal Society* 6: 1065–74.
65. Couzin ID (2009) Collective cognition in animal groups. *Trends in cognitive sciences* 13: 36–43.
66. Couzin ID, Krause J, James R, Ruxton GD, Franks NR (2010) Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* 218: 1–11.

Supporting Figures

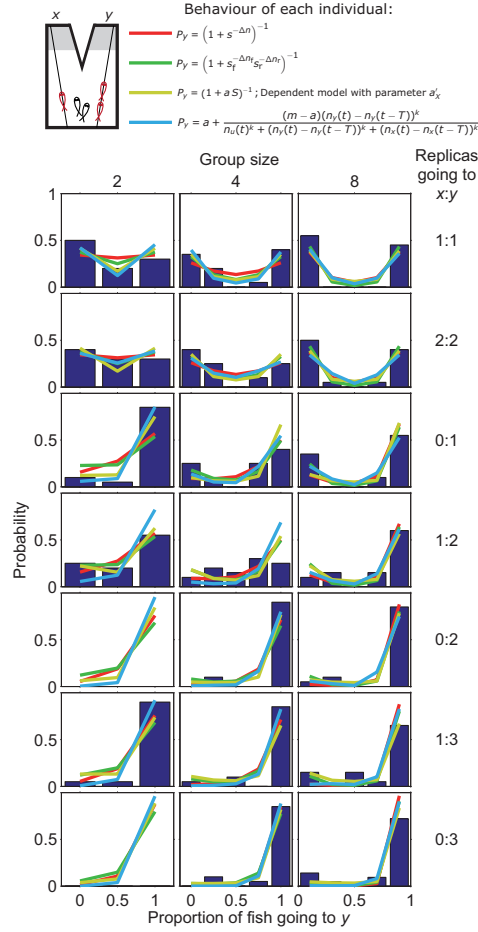


Figure S1. Comparison between different models for the symmetric set-up. Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histograms). Red line: results from our single-parameter model assuming independence in Eq. 17 in the main text ($s = 2.2$). Green line: Enhanced model assuming independence with different reliability for the replicas ($s_f = 3$, $s_r = 1.76$). Yellow line: Model including dependencies ($a'_x = 4.9$). Blue line: Empirical model presented in Ref. [42], using the parameters reported there. Different graphs correspond to different stickleback group sizes and different number of replicas going to x and y . According to Bayesian Information Criterion (BIC, see *Methods*), the best model is our model with dependencies (yellow line, logprob $L = -394$, and BIC weight $w = 0.996$). Second-best is the complicated version of the model without dependencies (green line, logprob $L = -396$, and BIC weight $w = 0.004$). Third-best is our one-parameter model assuming independence (red line, $L = -419$, $w = 3 \cdot 10^{-11}$). And last (but not far from the third one) the model from Ref. [42] (blue line, $L = -411$, $w = 5 \cdot 10^{-13}$). For the model from Ref. [42], L and w correspond to a re-optimization of the model as described in *Methods*, because using the parameters reported in [42] would perform worse).

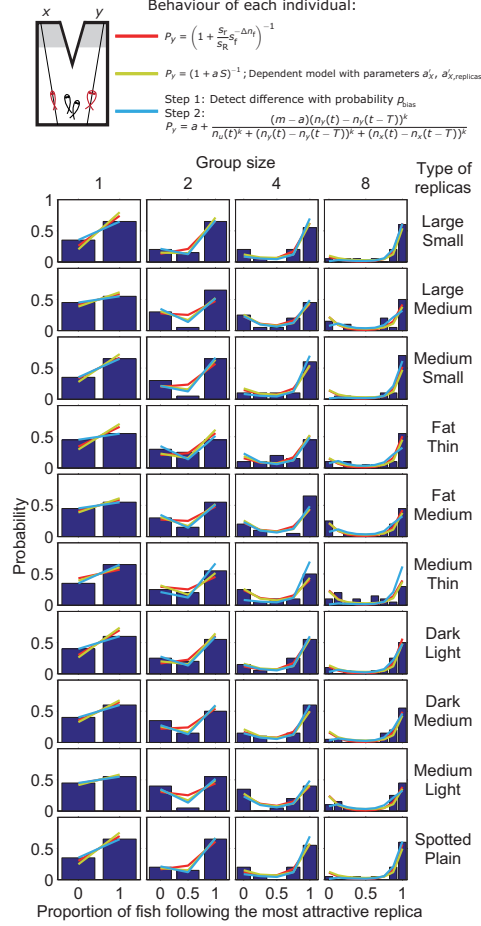


Figure S2. Comparison between different models for the condition with two different replicas. Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [43] (blue histograms). Red line: results from model in Eq. 20 in the main text ($s_f = 2.9$, $s_r/s_R = 0.35$, 0.7, 0.5, 0.52, 0.69, 0.75, 0.43, 0.55, 0.78, 0.43 for each row from top to bottom). Yellow line: Model including dependencies ($a'_X = 4.8$, $a'_{X,replicas} = 21.4, 11.8, 0.6, 9.9, 4.8, 0.9, 13, 8, 0.7, 14.5, 0.9$ for each type of replica (large, medium, small, etc.)). Blue line: Empirical model presented in Ref. [43], using the parameters reported there. Different graphs correspond to different stickleback group sizes and different types of replicas going to x and y . According to Bayesian Information Criterion (BIC, see *Methods*), our model neglecting dependencies gives the best representation of the data (red line, $\logprob L = -783$, and BIC weight $w = 0.9985$). Second-best is our model including dependencies, ($L = -788$, $w = 0.001$). Last, but near the second one, is the model from ref. [43] (blue line, $L = -781$, $w = 0.0005$). For the model from Ref. [43], these values of L and w correspond to a re-optimization of the model as described in *Methods*, because using the parameters reported in [43] would perform worse). The values of $\logprob (L)$ reported here do not include the data of the single-individual experiments (see *Methods*).

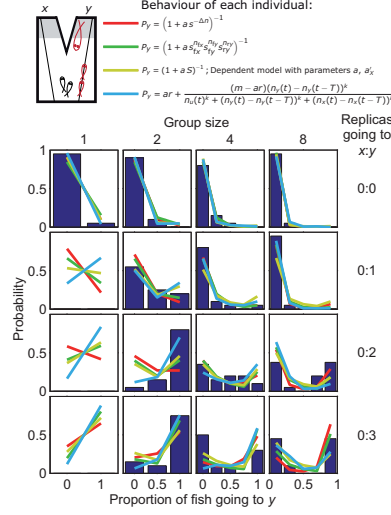


Figure S3. Comparison between different models in the asymmetrical set-up. Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histograms). Red line: results from model neglecting dependencies in Eq. 22 in the main text ($s = 2.6$, $a = 9.5$). Green line: Enhanced model neglecting dependencies with different reliability for the fish going to different locations and for the replicas ($a = 5.5$, $s_{tx} = 50$, $s_{ty} = 2/3$, $s_{ry} = 0.36$. s_{rx} has no effect because there are no replicas going to x). Yellow line: Two-parameter model including dependencies ($a = 9.94$, $a'_x = 8.66$). Blue line: Empirical model presented in Ref. [42], using the parameters reported there. Different graphs correspond to different stickleback group sizes and different number of replicas going to y . According to Bayesian Information Criterion (BIC, see *Methods*), the best two models are our complicated version neglecting dependencies (green line, logprob $L = -225$, and BIC weight $w = 0.52$) and our two-parameter model including dependencies (yellow line, $L = -231$, $w = 0.38$). Next (but very near) is our simplified model (red line, $L = -232$, $w = 0.098$). And last (and significantly worse) the model from Ref. [42] (blue line, $L = -234$, $w = 2.5 \cdot 10^{-6}$). For the model from Ref. [42], the values of L and w correspond to a re-optimization of the model as described in *Methods*, because using the parameters reported in [42] would perform worse. In two of the graphs for group size 1 that there are no data the prediction of the model from Ref. [42] and our model (especially the simplest version) are opposite. It might be that the results changed completely, depending on the results of these graphs, were the experiments performed. But we found that this is not the case: We performed simulations, adding experimental data in these two graphs. Even in the extreme case that the fabricated results matched exactly the predictions of the model in Ref. [42], BIC would still favour two of our models (we would get $L = -254$, $w = 0.99$ for our model with dependence, $L = -252$, $w = 0.01$ for our complicated model neglecting dependence, $L = -268$, $w = 8 \cdot 10^{-7}$ for our simplified model neglecting dependence and $L = -258$, $w = 3 \cdot 10^{-6}$ for the model in [42]).

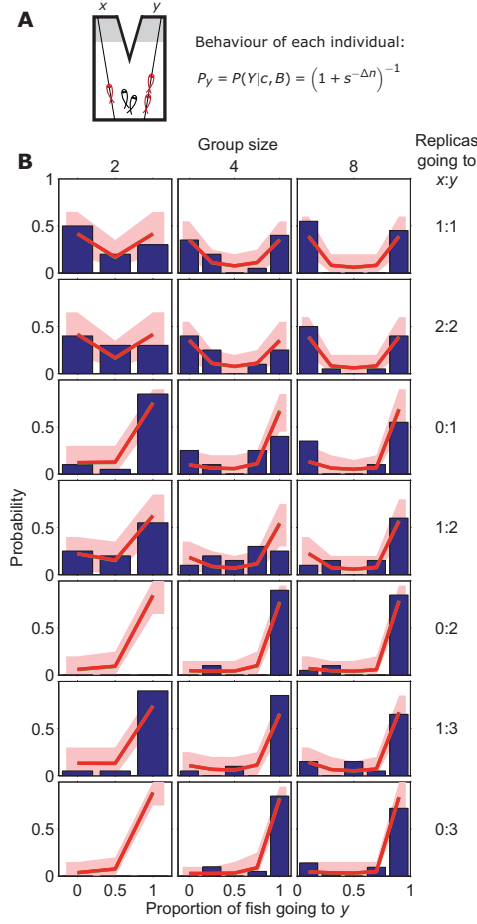


Figure S4. Comparison between model including dependencies and stickleback choices in symmetric set-up. (A) Schematic diagram of symmetric set-up with a group of sticklebacks (in black) choosing between two identical refugia and with different numbers of replica fish (in red) going to x and y . (B) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histogram) and results from the model that takes into account dependencies (red line using $a'_X = 4.9$; red region: 95% confidence interval; green line with $a'_X = 5$). Different graphs correspond to different stickleback group sizes and different number of replicas going to x and y .

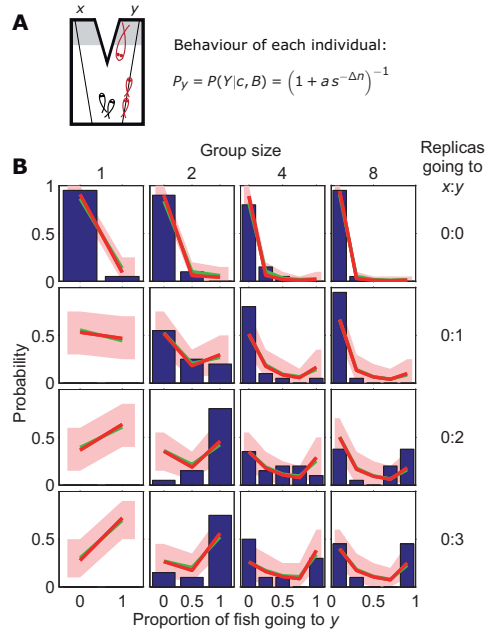


Figure S5. Comparison between model including dependencies and stickleback choices in asymmetric set-up. *A*) Schematic diagram of asymmetric set-up (predator at y , large fish depicted in red) with a group of sticklebacks (in black) choosing between two refugia, and replica fish (small fish depicted in red) going to y . (*B*) Experimentally measured statistics of final configurations of fish choices from 20 experimental repetitions [42] (blue histogram) and results from the model that takes into account the dependencies (red line using $a'_X = 8.7$, $a = 9.9$; red region: 95% confidence interval. Green line using $a'_X = 5$ and $a = 6.28$). Different graphs correspond to different stickleback group sizes and different number of replicas going to y .

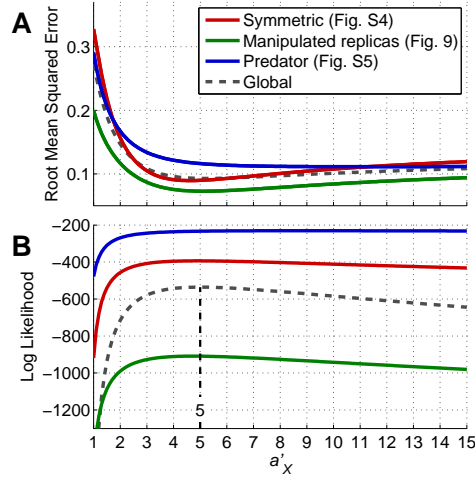


Figure S6. Goodness of fit of the model including dependencies for different values of a'_X . **Red:** Symmetric case (data in Fig. S4). **Green:** Case with different replicas at each side (data in Fig. 9). The parameters $a'_{X,\text{replicas}}$ are re-optimized for each value of a'_X . **Blue:** Asymmetric set-up with predator on one side (data in Fig. S5; Parameter a is re-optimized for each value of a'_X). (A) Root mean squared error between the data and the probabilities predicted by the model. Grey dashed line shows the mean RMSE for the three cases. The absolute values for each case depend on the shape of the data and are not comparable, only the trends and the position of the minima should be compared. (B) Logarithm of the probability that the data come from the model. The height of each curve depends on the number of data for each experiment, only the trend and the position of the maxima should be compared. Grey dashed line shows the sum of the three coloured lines, but shifted by 1000 so that it fits on the scale. The peak of this global probability indicates the value of a'_X that best fits the three datasets ($a'_X = 5$).

Supporting text: Model for more than 2 options

We present a derivation of the model for the more general case of M different options (instead of the 2 options used in the main text). We also discuss some particular cases that give simple expressions while still widely applicable.

Model for M options

Let M be the number of possible options, y_m , $m = 1 \dots M$. Each individual estimates the probability that each option is the best one, using its non-social information (C) and the behavior of the other individuals (B). So for one given option, say y_μ , we want to compute

$$P(Y_\mu|C, B), \quad (S1)$$

where Y_μ stands for ‘ y_μ is the best option’. We can compute the probability in Eq. S1 using Bayes’ theorem,

$$P(Y_\mu|C, B) = \frac{P(B|Y_\mu, C)P(Y_\mu|C)}{\sum_{m=1}^M P(B|Y_m, C)P(Y_m|C)}. \quad (S2)$$

Dividing numerator and denominator by the numerator, we get

$$P(Y_\mu|C, B) = \frac{1}{\sum_{m=1}^M a_{m\mu} S_{m\mu}}, \quad (S3)$$

where

$$a_{m\mu} = \frac{P(Y_m|C)}{P(Y_\mu|C)} \quad (S4)$$

contains only non-social information, and

$$S_{m\mu} = \frac{P(B|Y_m, C)}{P(B|Y_\mu, C)} \quad (S5)$$

contains the social information. Note that each term of the summation preserves the multiplicative relation between social and non-social information that was also apparent in Eq. 3 of the main text. There may be $M - 1$ independent non-social parameters $a_{m\mu}$ in the case that no two options have equal non-social information. But usually this will not be the case, and the number of independent non-social parameters will be lower.

Now we assume independence among behaviors (Eq. 6 in main text), and group all possible behaviors in L classes, $\{\beta_k\}_{k=1}^L$ (Eq. 7 in main text). These two assumptions transform Eq. S5 into

$$S_{m\mu} = \prod_{k=1}^L s_{k,m\mu}^{n_k}, \quad (S6)$$

where n_k is the number of individuals performing behavior β_k , and

$$s_{k,m\mu} = \frac{P(\beta_k|Y_m, C)}{P(\beta_k|Y_\mu, C)} \quad (S7)$$

are the reliability parameters for behavior β_k with respect to options y_m and y_μ . There may be up to $L(M - 1)$ independent reliability parameters but usually they will not be all independent.

In summary, from Equations S3 and S6 we have that

$$P(Y_\mu|C, B) = \left(\sum_{m=1}^M a_{m\mu} \prod_{k=1}^L s_{k,m\mu}^{n_k} \right)^{-1}. \quad (S8)$$

This equation summarizes the general model applicable to any kind of experiment. In the following sections we consider two particular cases with a much simpler expression.

One basic reliability parameter

The general model in Eq. S8 depends in general on $L(M - 1)$ independent reliability parameters $s_{k,m\mu}$. Here we derive the model for a particular case in which there is only one reliability parameter, s .

First, we consider classes of behaviors (from now on we call them just ‘behaviors’) that simply consist of choosing a given option. If for example the options are different places, behaviors would be going to each of those places. Therefore, the number of possible behaviors is the same as the number of options, $L = M$. We use the convention that β_j is ‘choosing option y_j ’. Note that when a behavior is not informative (i.e. its reliability parameter is 1) it has no impact on the model in Eq. S8. Therefore, considering this set of behaviors is equivalent to assuming that all other behaviors have reliability parameter equal to 1.

We further assume that $P(\beta_k|Y_m, C)$ only depends on whether $k = m$ or $k \neq m$, so that

$$\begin{aligned} P(\beta_k|Y_k, C) &= P(\beta_l|Y_l, C) \\ P(\beta_k|Y_m, C) &= P(\beta_l|Y_p, C), \quad k \neq m, l \neq p \end{aligned} \quad (S9)$$

Note that $P(\beta_k|Y_k, C)$ is the probability that another individual makes the correct choice, and $P(\beta_k|Y_m, C)$ with $k \neq m$ is the probability that it makes a wrong choice. So this assumption means that the probability of making the correct choice is the same regardless of which option is actually the correct one. In the case of symmetric choices, in which non-social information C is the same for all options, this relation will hold automatically, not being an extra assumption. It is likely that it also holds for many asymmetric choices. For example, the results for the asymmetric set-up presented in the main text suggest that it holds in that case. We define

$$\begin{aligned} p_c &\equiv P(\beta_k|Y_k, C) \\ p_f &\equiv P(\beta_k|Y_m, C), \quad k \neq m. \end{aligned} \quad (\text{S10})$$

As it only matters whether the behavior matches the correct choice or not, there are only four distinct types of reliability parameters $s_{k,m\mu}$ (Eq. S7):

$$\begin{aligned} s_{k,kk} &= \frac{P(\beta_k|Y_k, C)}{P(\beta_k|Y_k, C)} = \frac{p_c}{p_c} = 1 \\ s_{k,ml} &= \frac{P(\beta_k|Y_m, C)}{P(\beta_k|Y_l, C)} = \frac{p_f}{p_f} = 1, \quad k \neq m, \quad k \neq l \\ s_{k,km} &= \frac{P(\beta_k|Y_k, C)}{P(\beta_k|Y_m, C)} = \frac{p_c}{p_f} = s, \quad k \neq m \\ s_{k,mk} &= \frac{P(\beta_k|Y_m, C)}{P(\beta_k|Y_k, C)} = \frac{p_f}{p_c} = \frac{1}{s}, \quad k \neq m, \end{aligned} \quad (\text{S11})$$

where

$$s \equiv \frac{p_c}{p_f} \quad (\text{S12})$$

is the basic reliability parameter, equal to the probability that another individual makes the correct choice over the probability that it makes a mistake, for any behavior and for any individual. We regroup the terms in Eq. S8 so that it reflects the different types of $s_{k,m\mu}$ (Eq. S11), and get

$$P(Y_\mu|C, B) = \left(\sum_{m=1}^M a_{m\mu} s_{m,m\mu}^{n_m} s_{\mu,m\mu}^{n_\mu} \prod_{\substack{k=1 \\ k \neq m \\ k \neq \mu}}^L s_{k,m\mu}^{n_k} \right)^{-1}. \quad (\text{S13})$$

Using the relations in Eq. S11 we have that

$$P(Y_\mu|C, B) = \left(\sum_{m=1}^M a_{m\mu} s^{-(n_\mu - n_m)} \right)^{-1}. \quad (\text{S14})$$

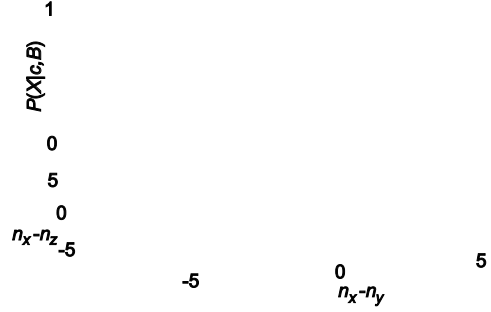


Figure S7. Probability of choosing one of the options for the 3-choice symmetric case.

Note that the term $m = \mu$ is always equal to 1, so Eq. S14 is identical to

$$P(Y_\mu|C, B) = \left(1 + \sum_{\substack{m=1 \\ m \neq \mu}}^M a_{m\mu} s^{-(n_\mu - n_m)} \right)^{-1}, \quad (\text{S15})$$

that has the same structure as the equations presented in the main text.

Symmetric case

In the special case that all options are indistinguishable using non-social information alone (symmetric case), all non-social parameters $a_{m\mu}$ are equal 1 and Eq. S15 becomes

$$P(Y_\mu|C, B) = \left(1 + \sum_{\substack{m=1 \\ m \neq \mu}}^M s^{-(n_\mu - n_m)} \right)^{-1}. \quad (\text{S16})$$

We recall that in this case Eq. S9 holds automatically, not being an extra assumption.

In the particular case of 3 options, x, y, z , we have

$$P(X|C, B) = \left(1 + s^{-(n_x - n_y)} + s^{-(n_x - n_z)} \right)^{-1}, \quad (\text{S17})$$

and the corresponding expressions for $P(Y|C, B)$ and $P(Z|C, B)$. Fig. S7 shows $P(X|C, B)$ in terms of its two effective variables, $n_x - n_y$ and $n_x - n_z$ (Eq. S17).